PCT

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(54) Title: ENHANCING PERCEPTUAL PERFORMANCE OF SBR AND RELATED HFR CODING METHODS BY ADAPTIVE NOISE-FLOOR ADDITION AND NOISE SUBSTITUTION LIMITING
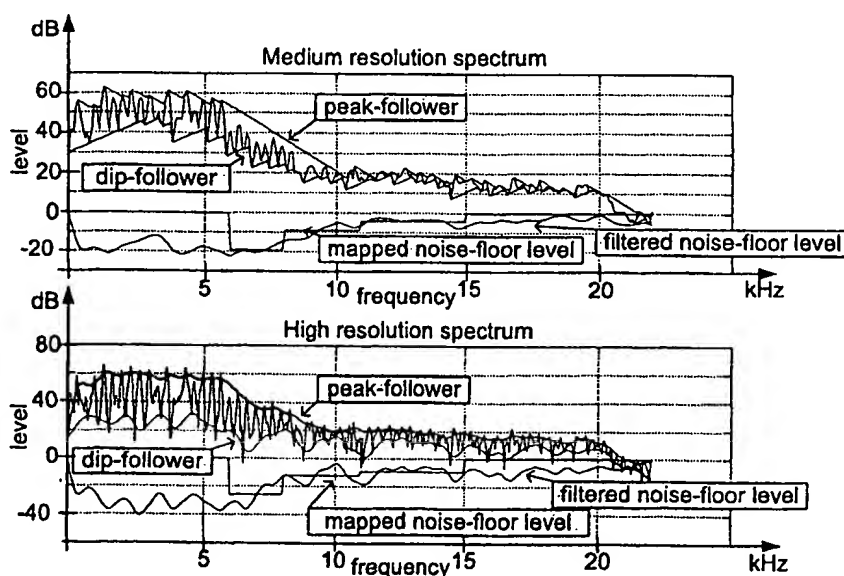
(57) Abstract

The present proposes new methods and an apparatus for enhancement of source coding systems utilising high frequency reconstruction (HFR). It addresses the problem of insufficient noise contents in a reconstructed highband, by Adaptive Noise-floor Addition. It also introduces new methods for enhanced performance by means of limiting unwanted noise, interpolation and smoothing of envelope adjustment amplification factors. The present invention is applicable to both speech coding and natural audio coding systems.

ENHANCING PERCEPTUAL PERFORMANCE OF SBR AND RELATED HFR CODING METHODS BY ADAPTIVE NOISE-FLOOR ADDITION AND NOISE SUBSTITUTION LIMITING

5

TECHNICAL FIELD

The present invention relates to source coding systems utilising high frequency reconstruction (HFR) such as Spectral Band Replication, SBR [WO 98/57436] or related methods. It improves performance of both high quality methods (SBR), as well as low quality copy-up methods [U.S. Pat. 5,127,054]. It is applicable to both speech coding

10 and natural audio coding systems. Furthermore, the invention can beneficially be used with natural audio codecs with- or without high-frequency reconstruction, to reduce the audible effect of frequency bands shut-down usually occurring under low bitrate conditions, by applying Adaptive Noise-flbor Addition.

BACKGROUND OF THE INVENTION

15 The presence of stochastic signal components is an important property of many musical instruments, as well as the human voice. Reproduction of these noise components, which usually are mixed with other signal components, is crucial if the signal is to be perceived as natural sounding. In high-frequency reconstruction it is, under certain conditions, imperative to add noise to the reconstructed high-band in order to achieve noise contents similar to the original. This necessity originates from the fact that most harmonic sounds, from for instance reed or bow

20 instruments, have a higher relative noise level in the high frequency region compared to the low frequency region. Furthermore, harmonic sounds sometimes occur together with a high frequency noise resulting in a signal with no similarity between noise levels of the highband and the low band. In either case, a frequency transposition, i.e. high quality SBR, as well as any low quality copy-up-process will occasionally suffer from lack of noise in the replicated highband. Even further, a high frequency reconstruction process usually comprises some sort of envelope adjustment,

25 where it is desirable to avoid unwanted noise substitution for harmonics. It is thus essential to be able to add and control noise levels in the high frequency regeneration process at the decoder.

Under low bitrate conditions natural audio codecs commonly display severe shut down of frequency bands. This is performed on a frame to frame basis resulting in spectral holes that can appear in an arbitrary fashion over the entire coded frequency range. This can cause audible artifacts. The effect of this can be alleviated by Adaptive Noise-floor

30 Addition.

Some prior art audio coding systems include means to recreate noise components at the decoder. This permits the encoder to omit noise components in the coding process, thus making it more efficient. However, for such methods

35 to be successful, the noise excluded in the encoding process by the encoder must not contain other signal components. This hard decision based noise coding scheme results in a relatively low duty cycle since most noise

components are usually mixed, in time and/or frequency, with other signal components. Furthermore it does not by any means solve the problem of insufficient noise contents in reconstructed high frequency bands.

5   SUMMARY OF THE INVENTION

The present invention addresses the problem of insufficient noise contents in a regenerated highband, and spectral holes due to frequency bands shut-down under low-bitrate conditions, by adaptively adding a noise-floor. It also prevents unwanted noise substitution for harmonics. This is performed by means of a noise-floor level estimation in the encoder, and adaptive noise-floor addition and unwanted noise substitution limiting at the decoder.

10

The Adaptive Noise-floor Addition and the Noise Substitution Limiting method comprise the following steps:
-    At an encoder, estimating the noise-floor level of an original signal, using dip- and peak-followers applied to a spectral representation of the original signal;
-    At an encoder mapping the noise-floor level to several frequency bands, or representing it using LPC or any other polynomial representation;

15

-    At an encoder or decoder, smoothing the noise-floor level in time and/or frequency;
-    At a decoder, shaping random noise in accordance to a spectral envelope representation of the original signal, and adjusting the noise in accordance to the noise-floor level estimated in the encoder;
-    At a decoder, smoothing the noise level in time and/or frequency;

20   -   Adding the noise-floor to the high-frequency reconstructed signal, either in the regenerated high-band, or in the shut-down frequency bands.
-    At a decoder, adjusting the spectral envelope of the high-frequency reconstructed signal using limiting of the envelope adjustment amplification factors.
-    At a decoder, using interpolation of the received spectral envelope, for increased frequency resolution, and thus

25   improved performance of the limiter.
-    At a decoder, applying smoothing to the envelope adjustment amplification factors.
-    At a decoder generating a high-frequency reconstructed signal which is the sum of several high-frequency reconstructed signals, originating from different lowband frequency ranges, and analysing the lowband to provide control data to the summation.

30

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will now be described by way of illustrative examples, not limiting the scope or spirit of the invention, with reference to the accompanying drawings, in which:

35   Fig. 1 illustrates the peak- and dip-follower applied to a high- and medium-resolution spectrum, and the mapping of the noise-floor to frequency bands, according to the present invention;
Fig. 2 illustrates the noise-floor with smoothing in time and frequency, according to the present invention;
Fig. 3 illustrates the spectrum of an original input signal;

Fig. 4 illustrates the spectrum of the output signal from a SBR process without Adaptive Noise-floor Addition;

Fig. 5 illustrates the spectrum of the output signal with SBR and Adaptive Noise-floor Addition, according to the present invention;

Fig. 6 illustrates the amplification factors for the spectral envelope adjustment filterbank, according to the present invention;

Fig. 7 illustrates the smoothing of amplification factors in the spectral envelope adjustment filterbank, according to the present invention;

Fig. 8 illustrates a possible implementation of the present invention, in a source coding system on the encoder side;

Fig. 9 illustrates a possible implementation of the present invention, in a source coding system on the decoder side.

DESCRIPTION OF PREFERRED EMBODIMENTS

The below-described embodiments are merely illustrative for the principles of the present invention for improvement of high frequency reconstruction systems. It is understood that modifications and variations of the arrangements and the details described herein will be apparent to others skilled in the art. It is the intent, therefore, to be limited only by the scope of the impending patent claims and not by the specific details presented by way of description and explanation of the embodiments herein.

Noise-floor level estimation

When analysing an audio signal spectrum with sufficient frequency resolution, formants, single sinusodials etc. are clearly visible, this is hereinafter referred to as the fine structured spectral envelope. However, if a low resolution is used, no fine details can be observed, this is hereinafter referred to as the coarse structured spectral envelope. The level of the noise-floor, albeit it is not necessarily noise by definition, as used throughout the present invention, refers to the ratio between a coarse structured spectral envelope interpolated along the local minimum points in the high resolution spectrum, and a coarse structured spectral envelope interpolated along the local maximum points in the high resolution spectrum. This measurement is obtained by computing a high resolution FFT for the signal segment, and applying a peak- and dip-follower, Fig. 1. The noise-floor level is then computed as the difference between the peak- and the dip-follower. With appropriate smoothing of this signal in time and frequency, a noise-floor level measure is obtained. The peak follower function and the dip follower function can be described according to eq. 1 and eq. 2,

$$Y_{peak}(X(k)) = \max(Y(X(k-1)) - T, X(k)) \quad \forall \quad 1 \le k \le \frac{fftSize}{2} \qquad \text{eq. 1}$$

$$Y_{dip}(X(k)) = \min(Y(X(k-1)) + T, X(k)) \quad \forall \quad 1 \le k \le \frac{fftSize}{2} \qquad \text{eq. 2}$$

where T is the decay factor, and $X(k)$ is the logarithmic absolute value of the spectrum at line k. The pair is calculated for two different FFT sizes, one high resolution and one medium resolution, in order to get a good estimate during vibratos and quasi-stationary sounds. The peak- and dip-followers applied to the high resolution FFT are LP-filtered

in order to discard extreme values. After obtaining the two noise-floor level estimates, the largest is chosen. In one implementation of the present invention the noise-floor level values are mapped to multiple frequency bands, however, other mappings could also be used e.g. curve fitting polynomials or LPC coefficients. It should be pointed out that several different approaches could be used when determining the noise contents in an audio signal. However it is, as described above, one objective of this invention, to estimate the difference between local minima and maxima in a high-resolution spectrum, albeit this is not necessarily an accurate measurement of the true noise-level. Other possible methods are linear prediction, autocorrelation etc, these are commonly used in hard decision noise/no noise algorithms ["Improving Audio Codecs by Noise Substitution" D. Schultz, JAES, Vol. 44, No. 7/8, 1996]. Although these methods strive to measure the amount of true noise in a signal, they are applicable for measuring a noise-floor-level as defined in the present invention, albeit not giving equally good results as the method outlined above. It is also possible to use an analysis by synthesis approach, i.e. having a decoder in the encoder and in this manner assessing a correct value of the amount of adaptive noise required.

Adaptive Noise-floor Addition

In order to apply the adaptive noise-floor, a spectral envelope representation of the signal must be available. This can be linear PCM values for filterbank implementations or an LPC representation. The noise-floor is shaped according to this envelope prior to adjusting it to correct levels, according to the values received by the decoder. It is also possible to adjust the levels with an additional offset given in the decoder.

In one decoder implementation of the present invention, the received noise-floor levels are compared to an upper limit given in the decoder, mapped to several filterbank channels and subsequently smoothed by LP filtering in both time and frequency, Fig. 2. The replicated highband signal is adjusted in order to obtain the correct total signal level after adding the noise-floor to the signal. The adjustment factors and noise-floor energies are calculated according to eq. 3 and eq. 4.

$$noiseLevel(k,l) = sfb\_nrg(k,l) \cdot \frac{nf(k,l)}{1 + nf(k,l)} \qquad \text{eq. 3}$$

$$adjustFactor(k,l) = \sqrt{\frac{1}{1 + nf(k,l)}} \qquad \text{eq. 4}$$

where k indicates the frequency line, l the time index for each sub-band sample, $sfb\_nrg(k,l)$ is the envelope representation, and $nf(k,l)$ is the noise-floor level. When noise is generated with energy $noiseLevel(k,l)$ and the highband amplitude is adjusted with $adjustFactor(k,l)$ the added noise-floor and highband will have energy in accordance with $sfb\_nrg(k,l)$. An example of the output from the algorithm is displayed in Fig. 3-5. Fig. 3 shows the spectrum of an original signal containing a very pronounced formant structure in the low band, but much less pronounced in the highband. Processing this with SBR without Adaptive Noise-floor Addition yields a result according to Fig. 4. Here it is evident that although the formant structure of the replicated highband is correct, the noise-floor level is too low. The noise-floor level estimated and applied according to the invention yields the result of

Fig. 5, where the noise-floor superimposed on the replicated highband is displayed. The benefit of Adaptive Noise-floor Addition is here very obvious both visually and audibly.

Transposer gain adaptation

An ideal replication process, utilising multiple transposition factors, produces a large number of harmonic components, providing a harmonic density similar to that of the original. A method to select appropriate amplification-factors for the different harmonics is described below. Assume that the input signal is a harmonic series:

$$x(t) = \sum_{i=0}^{N-1} a_i \cos(2\pi f_i t).$$

eq. 5

A transposition by a factor two yields:

$$y(t) = \sum_{i=0}^{N-1} a_i \cos(2 \times 2\pi f_i t).$$

eq. 6

Clearly, every second harmonic in the transposed signal is missing. In order to increase the harmonic density, harmonics from higher order transpositions, $M=3,5$ etc, are added to the highband. To benefit the most of multiple harmonics, it is important to appropriately adjust their levels to avoid one harmonic dominating over another within an overlapping frequency range. A problem that arises when doing so, is how to handle the differences in signal level between the source ranges of the harmonics. These differences also tend to vary between programme material, which makes it difficult to use constant gain factors for the different harmonics. A method for level adjustment of the harmonics that takes the spectral distribution in the low band into account is here explained. The outputs from the transposers are fed through gain adjusters, added and sent to the envelope-adjustment filterbank. Also sent to this filterbank is the low band signal enabling spectral analysis of the same. In the present invention the signal-powers of the source ranges corresponding to the different transposition factors are assessed and the gains of the harmonics are adjusted accordingly. A more elaborate solution is to estimate the slope of the low band spectrum and compensate for this prior to the filterbank, using simple filter implementations, e.g. shelving filters. It is important to note that this procedure does not affect the equalisation functionality of the filterbank, and that the low band analysed by the filterbank is not re-synthesised by the same.

Noise Substitution Limiting

According to the above (eq. 5 and eq. 6), the replicated highband will occasionally contain holes in the spectrum. The envelope adjustment algorithm strives to make the spectral envelope of the regenerated highband similar to that of the original. Suppose the original signal has a high energy within a frequency band, and that the transposed signal displays a spectral hole within this frequency band. This implies, provided the amplification factors are allowed to assume arbitrary values, that a very high amplification factor will be applied to this frequency band, and noise or other unwanted signal components will be adjusted to the same energy as that of the original. This is referred to as unwanted noise substitution. Let

$$P_1 = [p_{11}, ..., p_{1N}] \qquad \text{eq. 7}$$

be the scale factors of the original signal at a given time, and

$$P_2 = [p_{21}, ..., p_{2N}] \qquad \text{eq. 8}$$

the corresponding scale factors of the transposed signal, where every element of the two vectors represents sub-band energy normalised in time and frequency. The required amplification factors for the spectral envelope adjustment filterbank is obtained as

$$G = [g_1, ..., g_N] = \left[\sqrt{\frac{p_{11}}{p_{21}}}, ..., \sqrt{\frac{p_{1N}}{p_{2N}}} \right]. \qquad \text{eq. 9}$$

By observing G it is trivial to determine the frequency bands with unwanted noise substitution, since these exhibit much higher amplification factors than the others. The unwanted noise substitution is thus easily avoided by applying a limiter to the amplification factors, i.e. allowing them to vary freely up to a certain limit, g_{max}. The amplification factors using the noise-limiter is obtained by

$$G_{lim} = [\min(g_1, g_{max}), ..., \min(g_N, g_{max})]. \qquad \text{eq. 10}$$

However, this expression only displays the basic principle of the noise-limiters. Since the spectral envelope of the transposed and the original signal might differ significantly in both level and slope, it is not feasible to use constant values for g_{max}. Instead, the average gain, defined as

$$G_{avg} = \sqrt{\frac{\sum_i P_{1i}}{\sum_i P_{2i}}}, \qquad \text{eq. 11}$$

is calculated and the amplification factors are allowed to exceed that by a certain amount. In order to take wide-band level variations into account, it is also possible to divide the two vectors P_1 and P_2 into different sub-vectors, and process them accordingly. In this manner, a very efficient noise limiter is obtained, without interfering with, or confining, the functionality of the level-adjustment of the sub-band signals containing useful information.

Interpolation

It is common in sub-band audio coders to group the channels of the analysis filterbank, when generating scale factors. The scale factors represent an estimate of the spectral density within the frequency band containing the grouped analysis filterbank channels. In order to obtain the lowest possible bit rate it is desirable to minimise the number of scale factors transmitted, which implies the usage of as large groups of filter channels as possible. Usually this is done by grouping the frequency bands according to a Bark-scale, thus exploiting the logarithmic frequency resolution of the human auditory system. It is possible in an SBR-decoder envelope adjustment filterbank, to group the channels identically to the grouping used during the scale factor calculation in the encoder. However, the adjustment filterbank can still operate on a filterbank channel basis, by interpolating values from the received scale factors. The simplest interpolation method is to assign every filterbank channel within the group used for the scale factor calculation, the value of the scale factor. The transposed signal is also analysed and a scale factor per filterbank channel is calculated. These scale factors and the interpolated ones, representing the original spectral

envelope, are used to calculate the amplification factors according to the above. There are two major advantages with this frequency domain interpolation scheme. The transposed signal usually has a sparser spectrum than the original. A spectral smoothing is thus beneficial and such is made more efficient when it operates on narrow frequency bands, compared to wide bands. In other words, the generated harmonics can be better isolated and controlled by the envelope adjustment filterbank. Furthermore, the performance of the noise limiter is improved since spectral holes can be better estimated and controlled with higher frequency resolution.

Smoothing

It is advantageous, after obtaining the appropriate amplification factors, to apply smoothing in time and frequency, in order to avoid aliasing and ringing in the adjusting filterbank as well as ripple in the amplification factors. Fig. 6 displays the amplification factors to be multiplied with the corresponding subband samples. The figure displays two high-resolution blocks followed by three low-resolution blocks and one high resolution block. It also shows the decreasing frequency resolution at higher frequencies. The sharpness of Fig. 6 is eliminated in Fig. 7 by filtering of the amplification factors in both time and frequency, for example by employing a weighted moving average. It is important however, to maintain the transient structure for the short blocks in time in order not to reduce the transient response of the replicated frequency range. Similarly, it is important not to filter the amplification factors for the high-resolution blocks excessively in order to maintain the formant structure of the replicated frequency range. In Fig. 9b the filtering is intentionally exaggerated for better visibility.

Practical implementations

The present invention can be implemented in both hardware chips and DSPs, for various kinds of systems, for storage or transmission of signals, analogue or digital, using arbitrary codecs. Fig. 8 and Fig. 9 shows a possible implementation of the present invention. Here the high-band reconstruction is done by means of Spectral Band Replication, SBR. In Fig.8 the encoder side is displayed. The analogue input signal is fed to the A/D converter 801, and to an arbitrary audio coder, 802, as well as the noise-floor level estimation unit 803, and an envelope extraction unit 804. The coded information is multiplexed into a serial bitstream, 805, and transmitted or stored. In Fig. 9 a typical decoder implementation is displayed. The serial bitstream is de-multiplexed, 901, and the envelope data is decoded, 902, i.e. the spectral envelope of the high-band and the noise-floor level. The de-multiplexed source coded signal is decoded using an arbitrary audio decoder, 903, and up-sampled 904. In the present implementation SBR-transposition is applied in unit 905. In this unit the different harmonics are amplified using the feedback information from the analysis filterbank, 908, according to the present invention. The noise-floor level data is sent to the Adaptive Noise-floor Addition unit, 906, where a noise-floor is generated. The spectral envelope data is interpolated, 907, the amplification factors are limited 909, and smoothed 910, according to the present invention. The reconstructed high-band is adjusted 911 and the adaptive noise is added. Finally, the signal is re-synthesised 912 and added to the delayed 913 low-band. The digital output is converted back to an analogue waveform 914.

CLAIMS

1. A method for enhancement of source coding systems using high-frequency reconstruction, where said source coding system comprises an encoder representing all operations performed prior to storage or transmission, and a
5   decoder representing all operations performed after storage or transmission, **characterised by**:
    at said encoder, estimating the noise-floor level of an original signal;
    at said decoder, shaping random noise in accordance to a spectral envelope representation, and adjusting said noise in accordance to said noise-floor level estimated in said encoder;
    at said decoder, adding said noise to the high-frequency reconstructed signal.
10

2. A method according to claim 1, **characterised in** that said noise-floor level is mapped to several frequency bands.

3. A method according to claim 1, **characterised in** that said noise-floor level is represented using LPC, or any other polynomial representation.
15

4. A method according to claim 1, **characterised in** that said noise-floor level is estimated using dip- and peak-followers applied to a spectral representation of said original signal.

5. A method according to claim 1, **characterised in** that said noise-floor level is smoothed in time and/or frequency.
20

6. A method according to claim 1, **characterised in** that the level of said noise generated in the decoder is smoothed in time and/or frequency.

7. A method according to claim 1, **characterised in** that the spectral envelope of said high-frequency reconstructed
25   signal is adjusted using limiting of the envelope adjustment amplification factors.

8. A method according to claim 1, **characterised in** that the spectral envelope of said high-frequency reconstructed signal is adjusted using interpolation.

30   9. A method according to claim 1, **characterised in** that the spectral envelope of said high-frequency reconstructed signal is adjusted using smoothing of the envelope adjustment amplification factors.

10. A method according to claim 1, **characterised in** that the high-frequency reconstruction generates a signal which is the sum of several high-frequency reconstructed signals, originating from different lowband frequency ranges, and
35   that an envelope adjustment device analyses said lowband and provides control data to said summation.

11. An apparatus for enhancement of source coding systems using high-frequency reconstruction, where said apparatus comprises an encoder, for encoding a signal to be decoded by a decoder, **characterised by**:

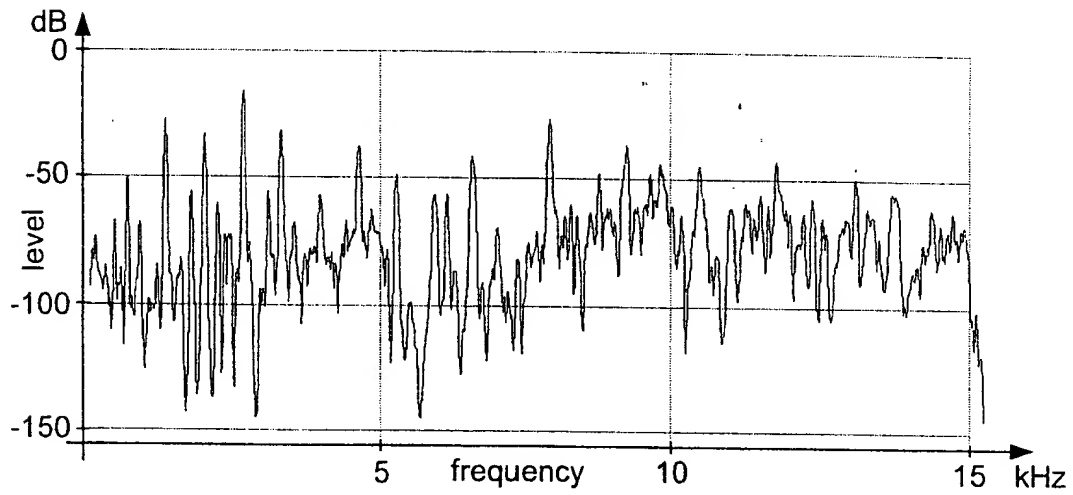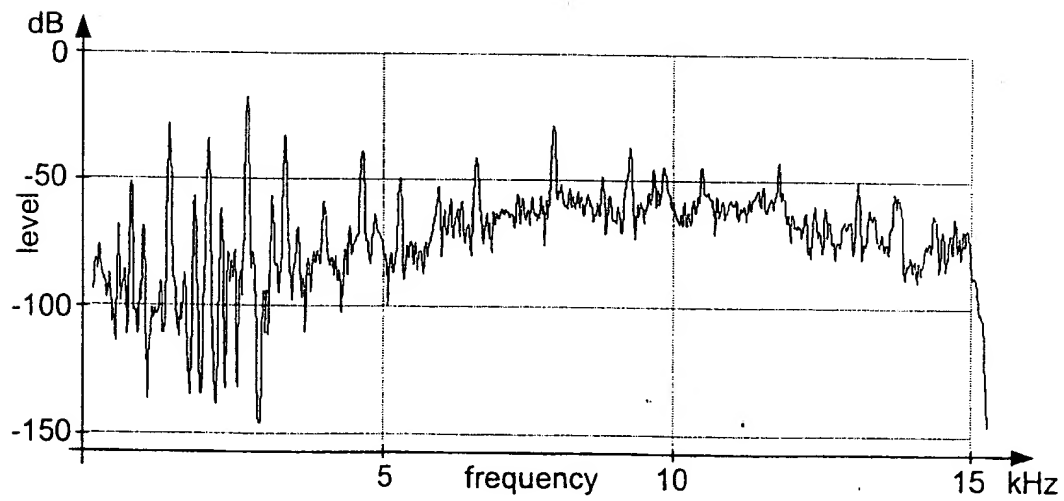means for estimating the noise-floor level of an original signal;

12. An apparatus for enhancement of source coding systems using high-frequency reconstruction, where said apparatus comprises a decoder, for decoding a coded signal encoded by an encoder, **characterised by**:
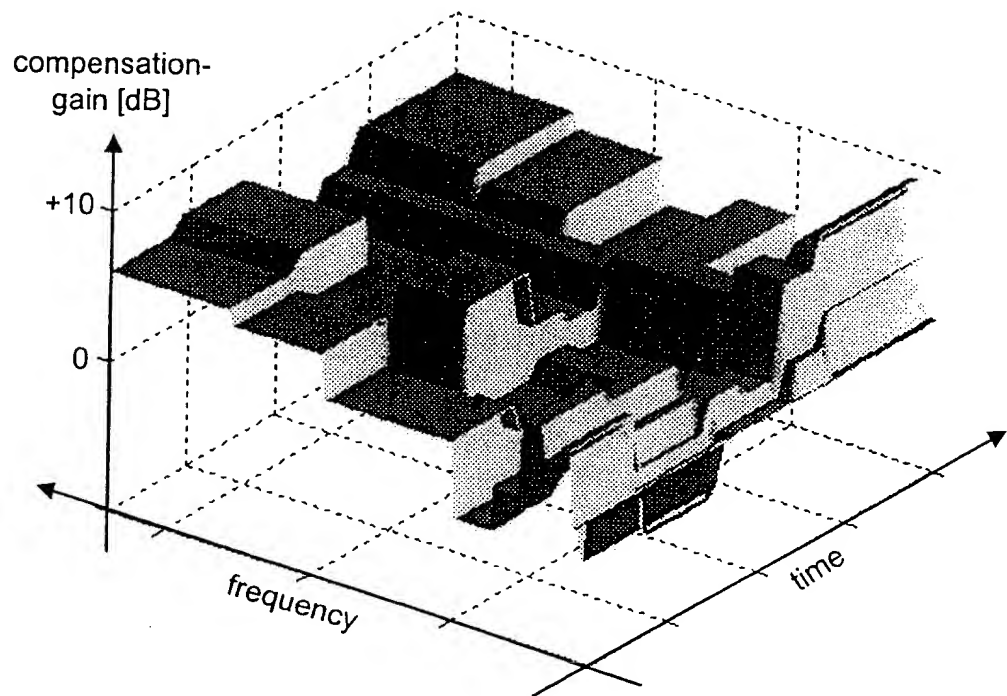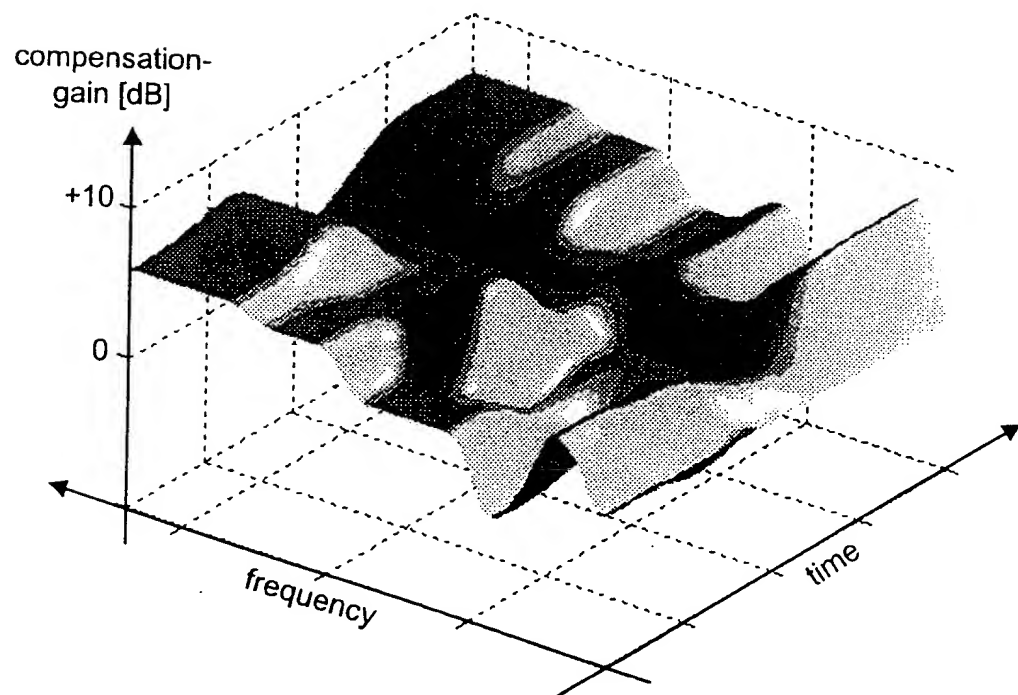
5   means for shaping random noise in accordance to a spectral envelope representation, and adjusting said noise in accordance to said noise-floor level estimated in said encoder;
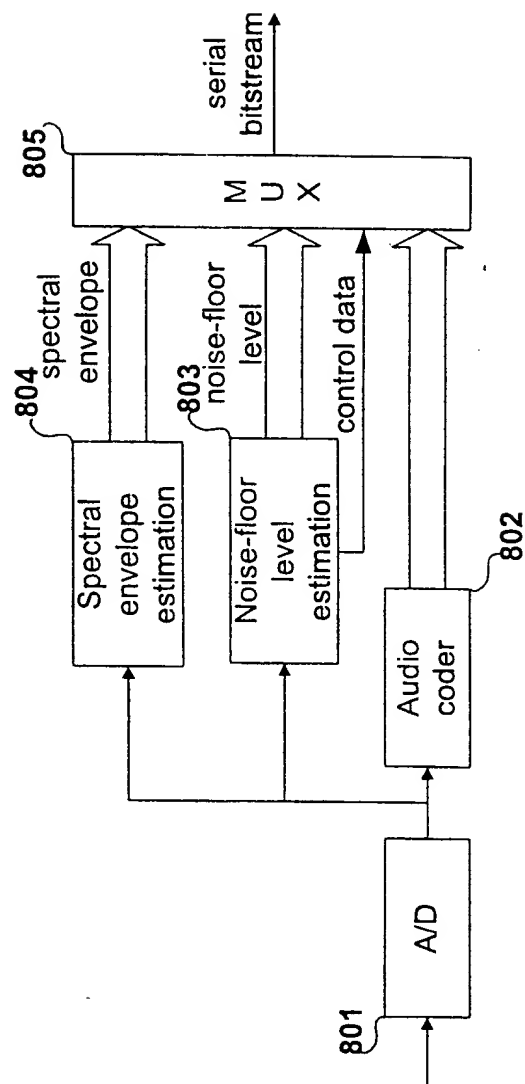
means for adding said noise to the high-frequency reconstructed signal.
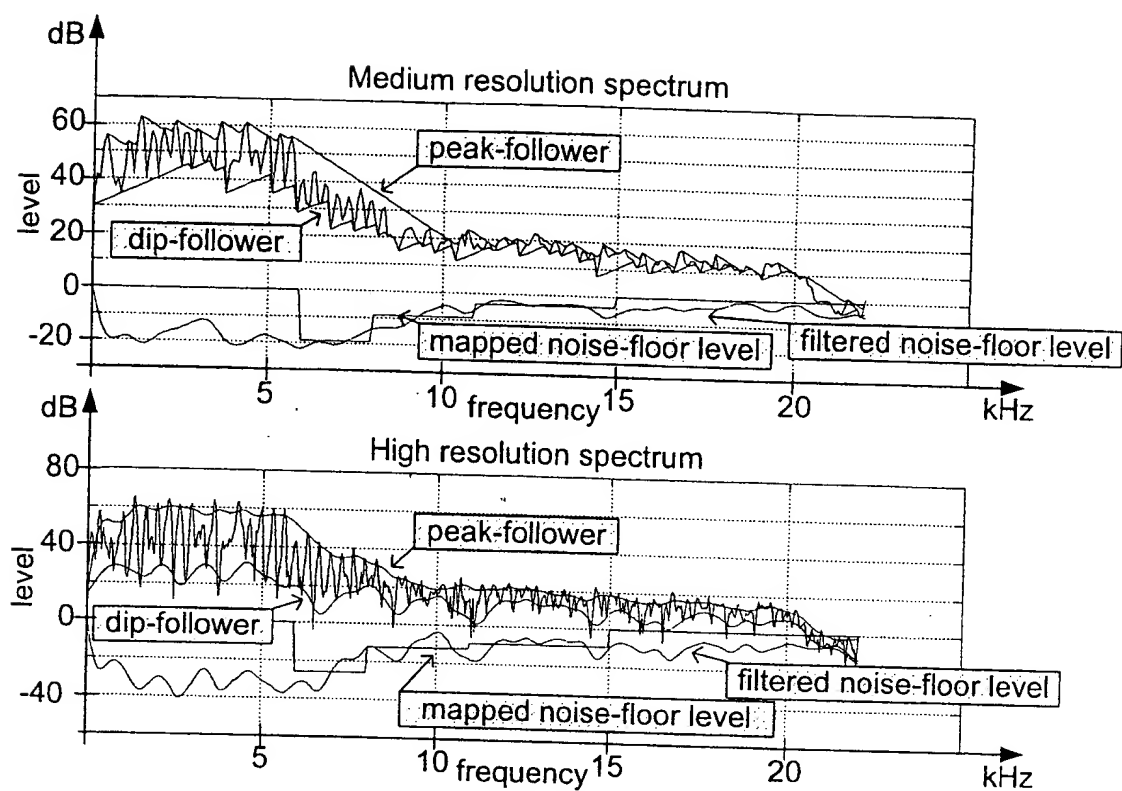
Fig. 1
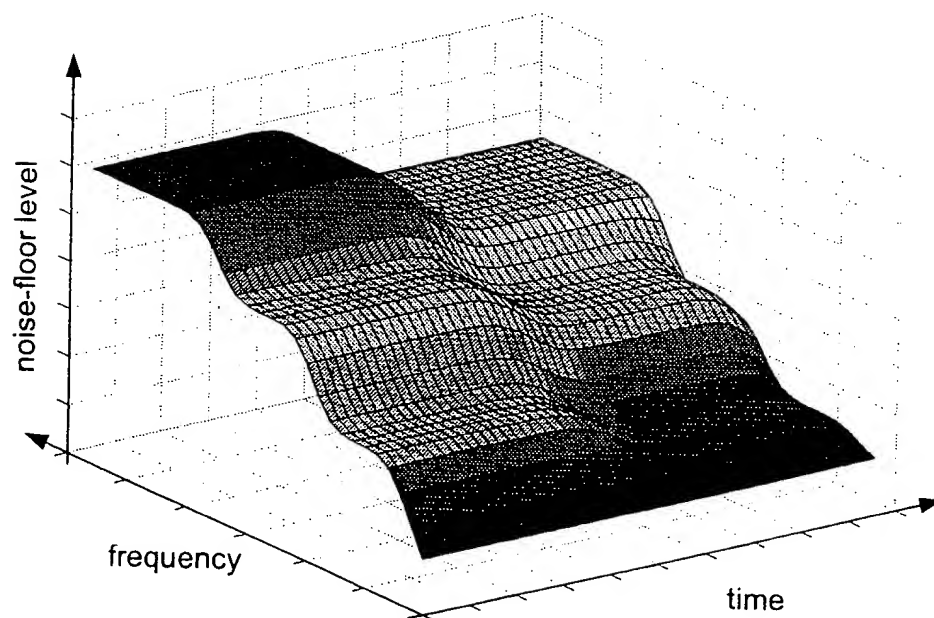
Fig. 2
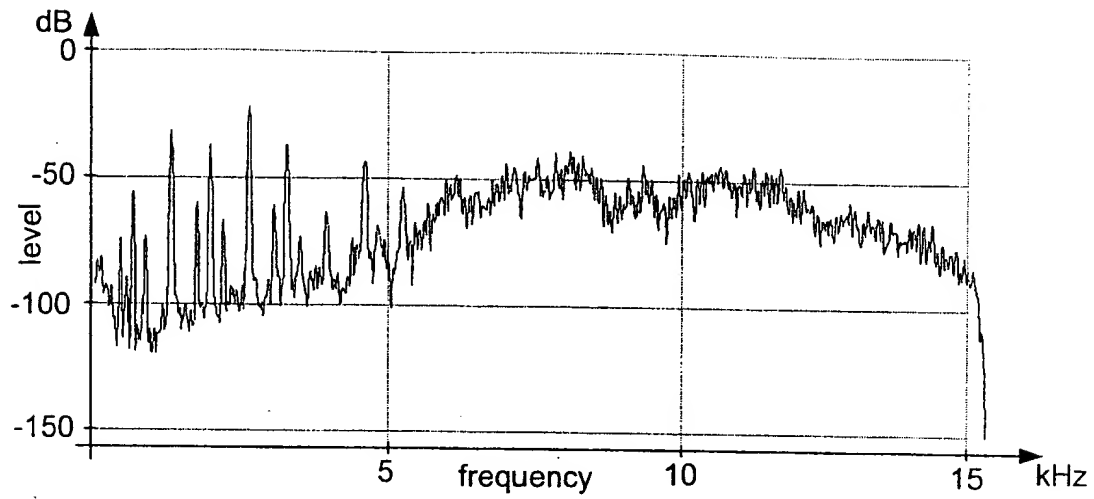
Fig. 3

Fig. 4

Fig. 5
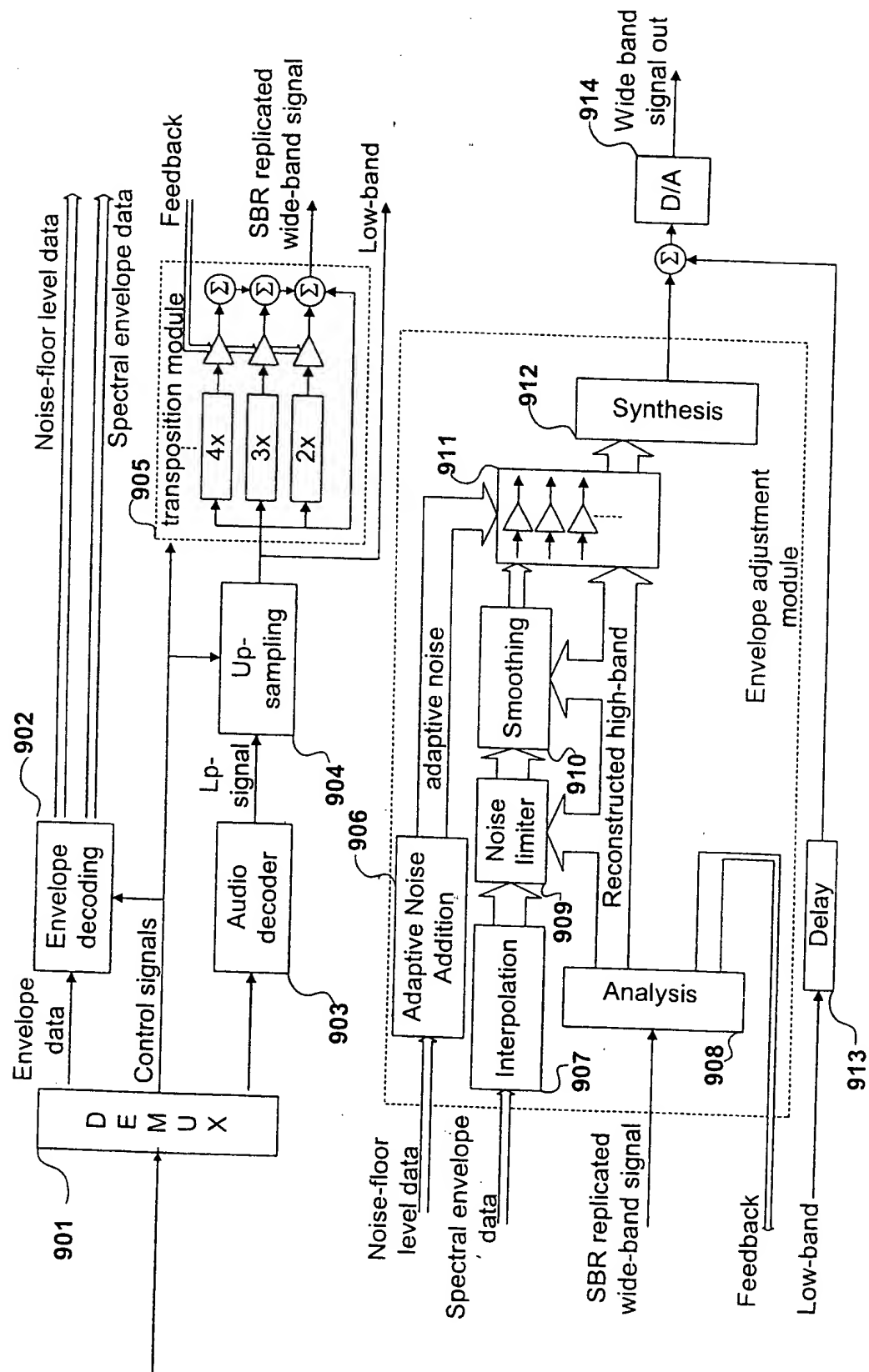
Fig. 6

Fig. 7

Fig. 8

Fig. 9